LPC

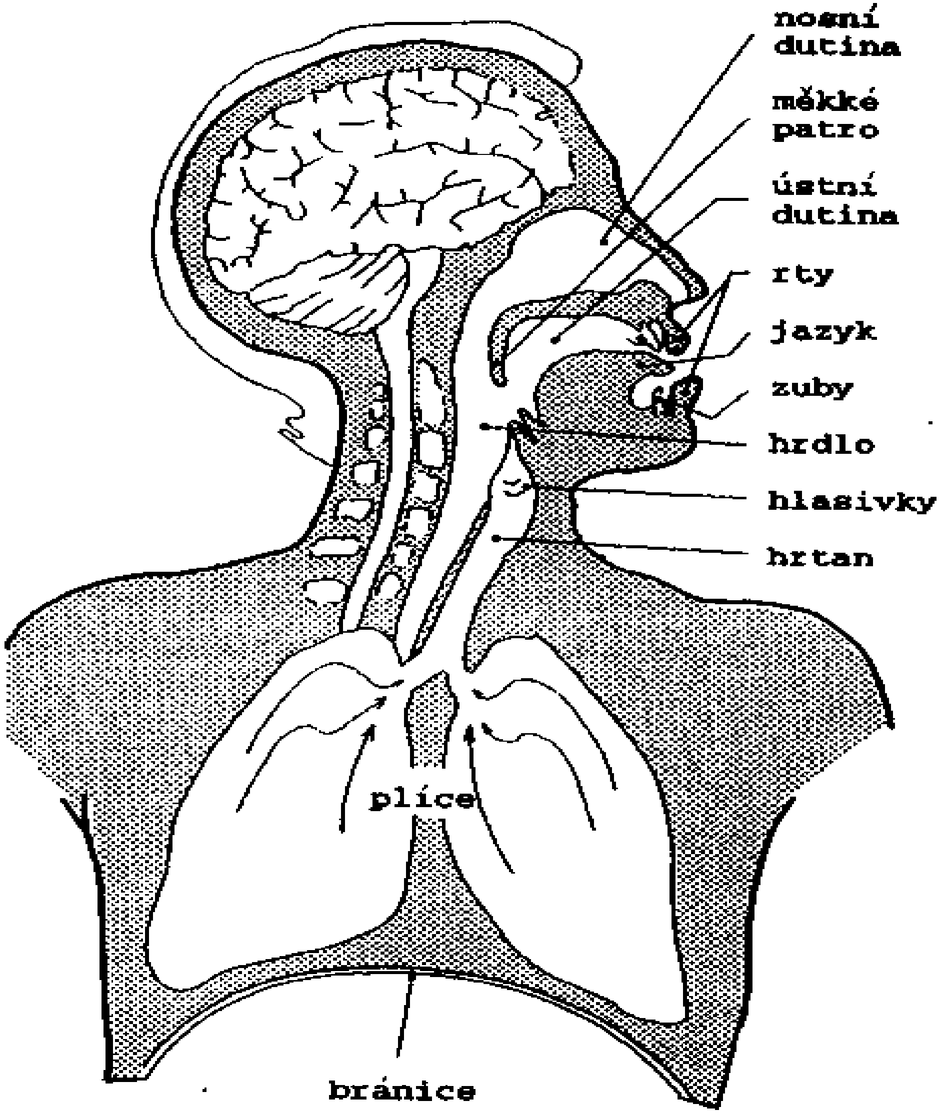
Jan Černocký ÚPGM FIT VUT Brno, `cernocky@fit.vutbr.cz`

FIT VUT Brno

Plán

- signálový model artikulačního traktu.
- proč lineární predikce.
- odhad koeficientů filtru (aneb posuňte si své signály!)
- Levinson-Durbin
- Spektrální hustota výkonu (PSD) pomocí LPC.
- Parametry odvozené z LPC.

Opakování – tvorba řeči a její model



Cíl: Chceme odhadnout parametry tohoto modelu. Tato přednáška se bude zabývat filtrem.

Model artikulačního traktu

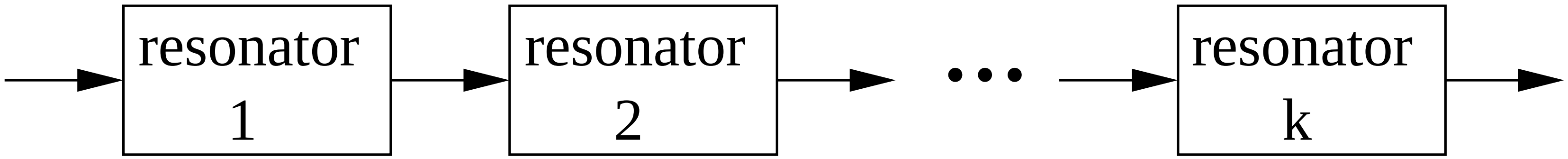
Hlasivky

dolní propusť 2. řádu, lomová frekvence okolo 100 Hz:

$$G(z) = \frac{1}{[1 - e^{-cT_s} z^{-1}]^2} \quad (1)$$

Hlasový trakt

kaskáda malých dvojpólových *rezonátorů* odpovídajících *formantům*.



Pro k formantů F_i s šířkami pásem B_i:

$$V(z) = \frac{1}{\prod_{i=1}^{K} [1 - 2e^{-\alpha_i T_s} \cos \beta_i T_s z^{-1} + e^{-2\alpha_i T_s} z^{-2}]} \qquad (2)$$

kde parametry α_i a β_i jsou určeny polohou a šířkou pásma formantů.

Model vyzařování zvuku

$$L(z) = 1 - z^{-1} \qquad (3)$$

což je horní propusť.

Dohromady . . .

$$
\begin{aligned}
H(z) &= G(z)V(z)L(z) = \\
&= \frac{1-z^{-1}}{(1-e^{-cT_s}z^{-1})^2\prod_{i=1}^{K}[1-2e^{-\alpha_i T_s}\cos\beta_i T_s z^{-1}+e^{-2\alpha_i T_s}z^{-2}]}
\end{aligned}
\tag{4}
$$

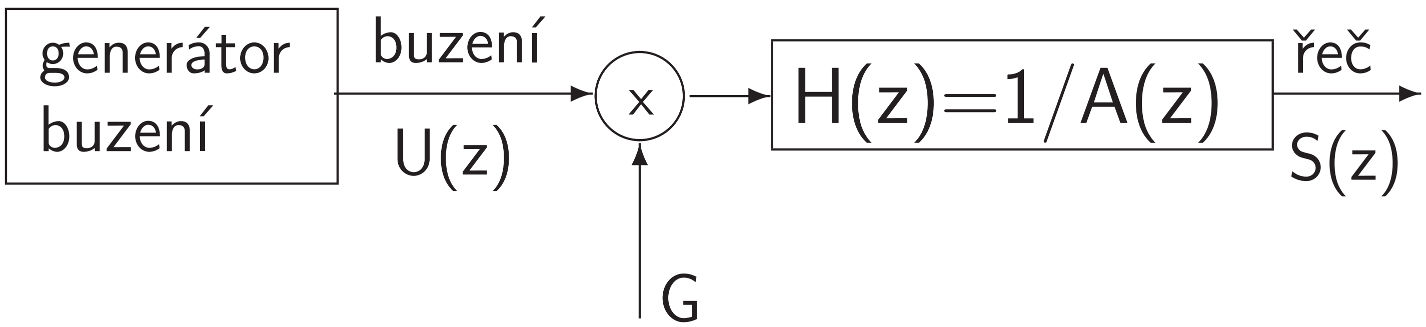
Člen $cT_s \rightarrow 0$, proto můžeme krátit čitatele i jmenovatele o jeden člen $1-z^{-1}$. Celkový model je tedy **celopólový** (obsahuje jen jmenovatele – čistý IIR filtr). Běžný zápis:

$$
H(z) = \frac{1}{1+\sum_{i=1}^{P} a_i z^{-i}} = \frac{1}{A(z)},
\tag{5}
$$

kde polynom $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_P z^{-P}$ má řád $P = 2k+1$ (k je počet formantů). Za užitečný počet pokládáme k=4 či 5, proto volíme často P=10 (pro F_s=8 kHz). Pro vyšší vzorkovací frekvence volíme P vyšší (např. 16), abychom postihli i vf. část spektra.

Určení parametrů modelu pomocí lineární predikce (LP)

Tvorba řeči s tímto filtrem:



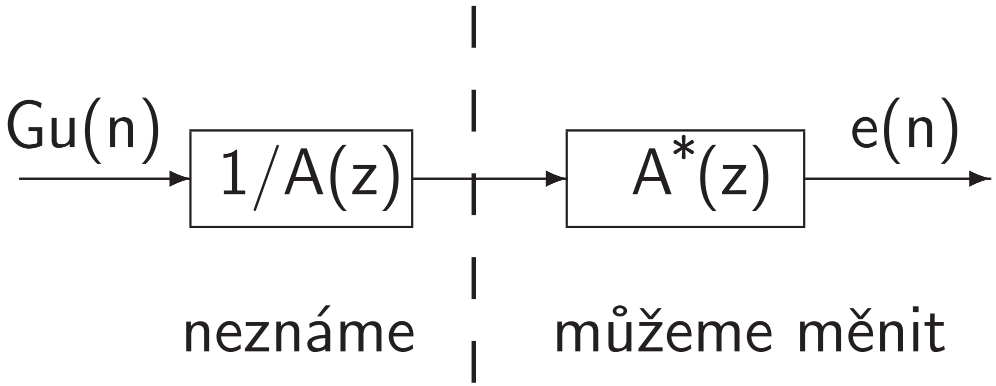
n-tý vzorek řeči je tedy dán:

$$s(n) = Gu(n) - \sum_{i=1}^{P} a_i s(n-i) \tag{6}$$

Parametry (koeficienty) filtru a_i jsou ovšem **neznámé** a musíme je **odhadnout**, odborně **identifikovat** (system identification).

Určení parametrů filtru

Můžeme zkonstruovat tzv. *inverzní filtr* $A^\star(z)$ s koeficienty α_i:



Ukazuje se, že v případě stacionárního signálu $s(n)$ jsou koeficienty a_i *identifikovány* pomocí koeficientů α_i, je-li *minimalizována energie signálu na výstupu* $e(n)$: $\mathcal{E}\{e^2(n)\}$.
"kroutíme parametry filtru tak dlouho, dokud není energie signálu na výstupu minimální. . . "

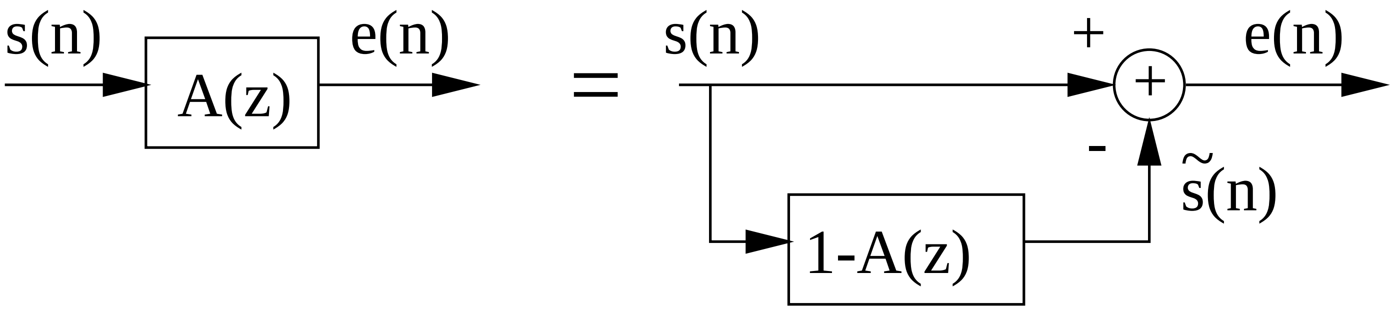
Proč "lineární predikce" ?

Předpokláme, že $\mathcal{E}\{e^2(n)\}$ je již minimalizována, tedy že $A^\star(z) = A(z)$ a budeme tedy používat pouze označení koeficientů a_i.

$A(z)$ můžeme (trochu podivně – jako když se levou rukou drbete za pravým uchem. . .) zapsat jako:

$$A(z) = 1 - [1 - A(z)] \tag{7}$$

a tedy:



Signál $\tilde{s}(n)$ je dán lineární kombinací několika předchozích vzorků, považujeme jej za *předpověď* skutečného vzorku $s(n)$:

$$\tilde{s}(n) = -\sum_{i=1}^{P} a_i s(n-i) \tag{8}$$

Chyba predikce je dána jako rozdíl skutečné a předpovězené hodnoty:

$$e(n) = s(n) - \tilde{s}(n) = s(n) - [-\sum_{i=1}^{P} a_i s(n-i)] = s(n) + \sum_{i=1}^{P} a_i s(n-i). \qquad (9)$$

Čím lepší predikce, tím menší chyba.

V rovině z:

$$E(z) = S(z)A(z) \qquad (10)$$

Výhody získání parametrů touto metodou:

- je-li $\alpha_i = a_i$, je chyba predikce rovna *buzení* (můžeme se tedy dostat ke vstupu do hlasového traktu bez skalpelu).
- určení koeficientů pomocí LP vede k soustavě snadno řešitelných lineárních rovnic.

Řešení

V této etapě řešení zatím záměrně nezmiňujeme, kolik vzorků vstupního signálu máme k dispozici, sumy jsou tedy zatím bez mezí. Nenormalizovaná energie chyby predikce je dána:

$$E = \sum_n e^2(n) \tag{11}$$

Tento výraz je třeba minimalizovat. Vyjádříme jej pomocí signálu $s(n)$ (známá veličina) a neznámých koeficientů a_i. Pro nalezení minima budeme parciálně derivovat podle každého a_i, derivace položíme rovny nule:

$$\frac{\delta}{\delta a_j}\left\{\sum_n [s(n) + \sum_{i=1}^{P} a_i s(n-i)]^2\right\} = 0 \tag{12}$$

$$\sum_n 2[s(n) + \sum_{i=1}^{P} a_i s(n-i)]s(n-j) = 0 \tag{13}$$

$$\sum_n s(n)s(n-j) + \sum_{i=1}^{P} a_i \sum_n s(n-i)s(n-j) = 0. \tag{14}$$

(15)

Označíme:

$$\sum_n s(n-i)s(n-j) = \phi(i,j), \tag{16}$$

pak

$$\sum_{i=1}^{P} a_i \phi(i,j) = -\phi(0,j) \quad \text{pro} \quad 1 \leq j \leq P \tag{17}$$

Což je soustava lineárních rovnic:

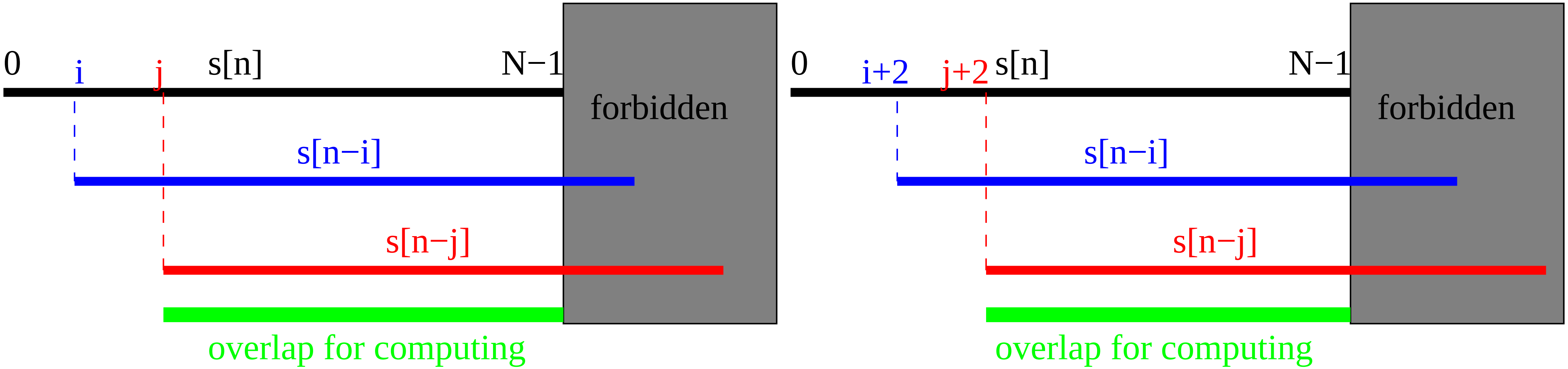
$$\begin{array}{ccc} \phi(1,1)a_1 + \phi(2,1)a_2 + & \cdots & +\phi(P,1)a_P = -\phi(0,1) \\ \phi(1,2)a_1 + \phi(2,2)a_2 + & \cdots & +\phi(P,2)a_P = -\phi(0,2) \\ & \vdots & \\ \phi(1,P)a_1 + \phi(2,P)a_2 + & \cdots & +\phi(P,P)a_P = -\phi(0,P), \end{array} \tag{18}$$

Výpočet $\phi(\cdot, \cdot)$

Koeficienty odhadujeme na rámci o délce N vzorků. Dvě metody lišící se v tom, jak nahlížíme na signál vně rámce (tedy pro vzorky $n < 0$ a $n > N - 1$):

Kovarianční metoda:

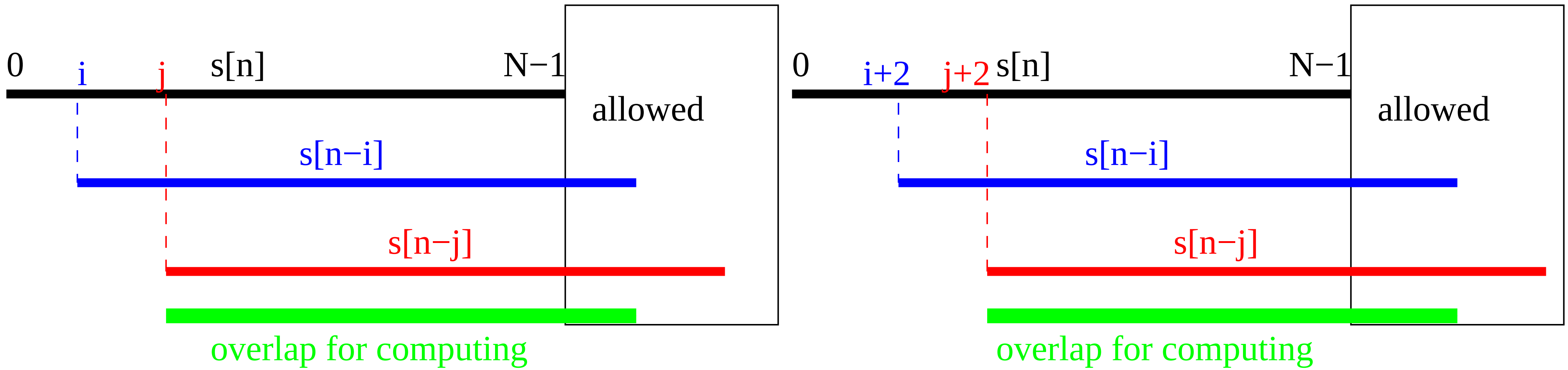
signál vně rámce je **neznámý**: se vzorky mimo $[0, N-1]$ nemohu počítat, ani když je tam signál zpožděn.



$\Rightarrow \phi(i,j)$ a $\phi(i+const, j+const)$ NEJSOU stejné (máme pokaždé jiný počet vzorků) - musí se řešit plná soustava lineárních rovnic. Složité, kovarianční metoda navíc vede k **nestabilnímu** filtru $1/A(z)$.

Korelační metoda:

signál vně rámce je považován za **známý, ale nulový** – mohu s ním počítat.



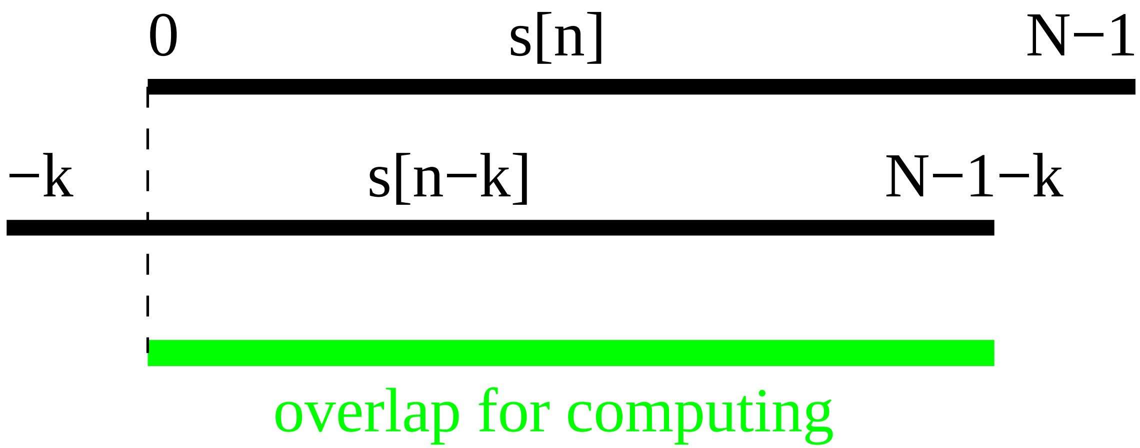
$\Rightarrow \phi(i,j)$ a $\phi(i+const, j+const)$ JSOU stejné (pokaždé stejné vzorky) - soustava rovnic se nám bude dobře počítat, protože na diagonálách budou stejné hodnoty: např. $\phi(2,1) = \phi(3,2) = \phi(4,3) = \cdots$

Proč jsou ϕ autokorelační koeficienty

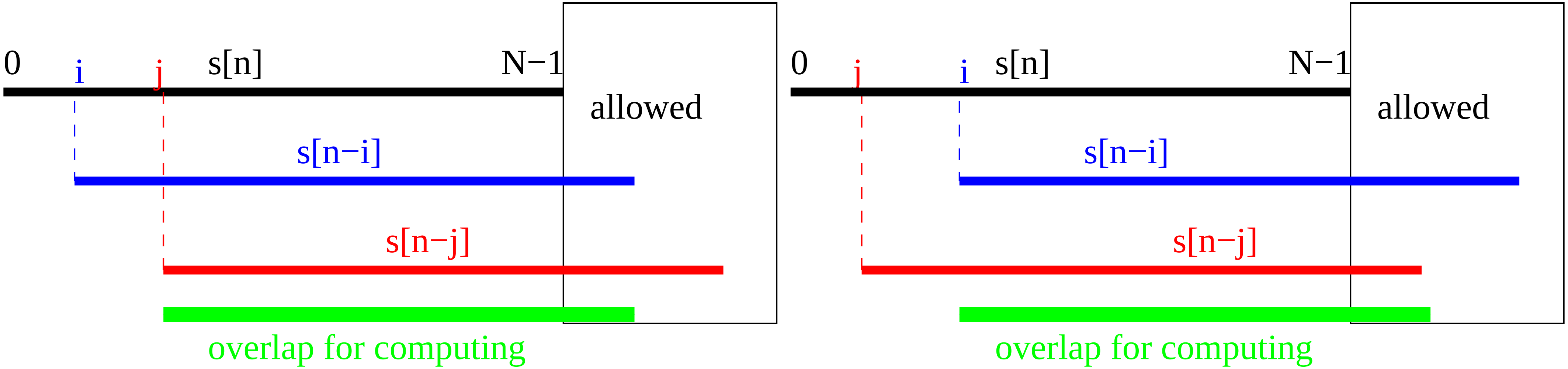
Odhad autokorelačních koeficientů (bez normalizace) pro signál o délce N pro kladné k, viz Signály a systémy, Náhodné procesy II.: `http://www.fit.vutbr.cz/~cernocky/sig`

$$R(k) = \sum_{n=0}^{N-1-k} s(n)s(n+k)$$

Korelační koeficienty "udávájí podobnost signálu samotného se sebou, když ho posuneme o k vzorků"



Situace pro $\phi(i,j)$ a $\phi(j,i)$:



$\Rightarrow$ pokaždé počítáme se stejnými vzorku $\Rightarrow$ oba dva jsou rovny autokorelačnímu koeficientu $R(|i-j|)$. To je fajn, protože matice bude navíc ještě symetrická. Matici, která je symetrická a má na diagonálách stejné prvky, se říká **Töplitzova**.

Výsledná soustava rovnic pro koeficienty $a_1 \dots a_P$

$$\begin{array}{lcr}
R(0)a_1 + R(1)a_2 + & \cdots & +R(P-1)a_P = -R(1) \\
R(1)a_1 + R(0)a_2 + & \cdots & +R(P-2)a_P = -R(2) \\
 & \vdots & \\
R(P-1)a_1 + R(P-2)a_2 + & \cdots & +R(0)a_P = -R(P),
\end{array} \quad (19)$$

Energie chyby predikce

Bez odvození . . . pomocí LPC můžeme dostat i následující vzoreček pro výpočet **nenormalizované** energie chyby predikce:

$$E = \sum_{n=0}^{N+P-1} e^2(n) = R(0) + \sum_{i=1}^{P} a_i R(i) \tag{20}$$

Pokud má budící signál *normovanou energii* rovnu 1 — např. bílý šum s rozptylem 1 nebo pulsy, kde $\frac{1}{N}\sum_{n=0}^{N-1} u^2(n) = 1$, pak, abychom dostali tutéž energii jako $s(n)$, musíme nastavit gain (zesílení) filtru na:

$$G^2 = \frac{E}{N} = \frac{1}{N}\left[R(0) + \sum_{i=1}^{P} a_i R(i)\right]. \tag{21}$$

. . . bude se nám hodit při kódování.

Levinson–Durbin

Jelikož je matice $\mathbf{R}$ symetrická a Töplitzova (všechny prvky na diagonálách jsou stejné), dá se k řešení soustavy 19 použít rychlý algoritmus Levinsona a Durbina:

$$E^{(0)} = R(0) \tag{22}$$

$$k_i = -\left[R(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j)\right] / E^{(i-1)} \tag{23}$$

$$a_i^{(i)} = k_i \tag{24}$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)} \quad \text{pro} \quad 1 \leq j \leq i-1 \tag{25}$$

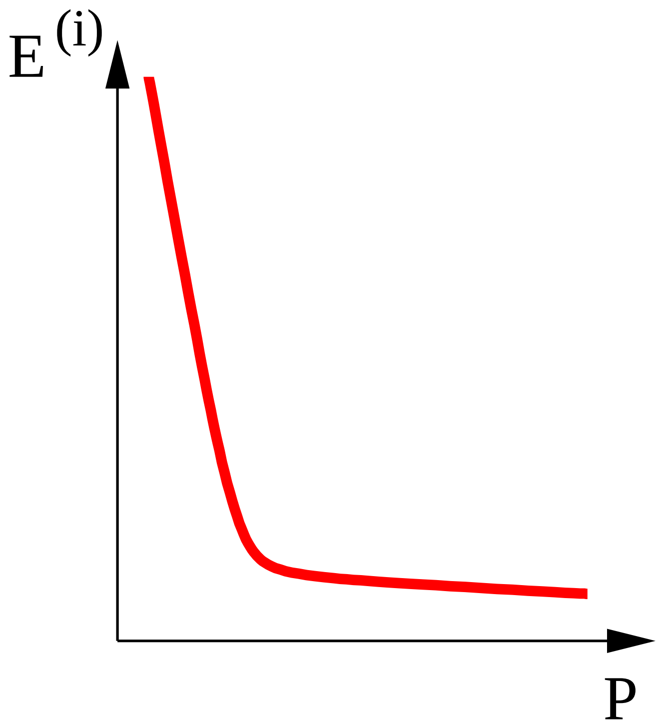
$$E^{(i)} = (1-k_i^2) E^{(i-1)} \tag{26}$$

Levinson-Durbin II.

- Postupně avyšujeme řád prediktoru (sloupce následující tabulky). $a_j^{(i)}$ je j-tý koeficient prediktoru řádu i:

$$
\begin{array}{ccccc}
a_1^{(1)} & a_1^{(2)} & a_1^{(3)} & \cdots & a_1^{(P)} \\
 & a_2^{(2)} & a_2^{(3)} & \cdots & a_2^{(P)} \\
 & & a_3^{(3)} & \cdots & a_3^{(P)} \\
 & & & \ddots & \vdots \\
 & & & & a_P^{(P)}
\end{array}
\tag{27}
$$

- podle průběhu energie chyby predikce $E(i)$ v závislosti na řádu prediktoru lze optimalizovat tento řád:



Zvyšování řádu prediktoru nad “lom” funkce již nepřináší téměř žádné zlepšení energie chyby.

Odhad spektrální hustoty výkonu (PSD) pomocí modelu LPC

Prozatím jsme PSD odhadovali pomocí DFT (obsahovala i "jemnou" složku způsobenou násobky frekvence základního tónu). PSD se dá ale odhadnout i pomocí frekvenční charakteristiky filtru $1/A(z)$:

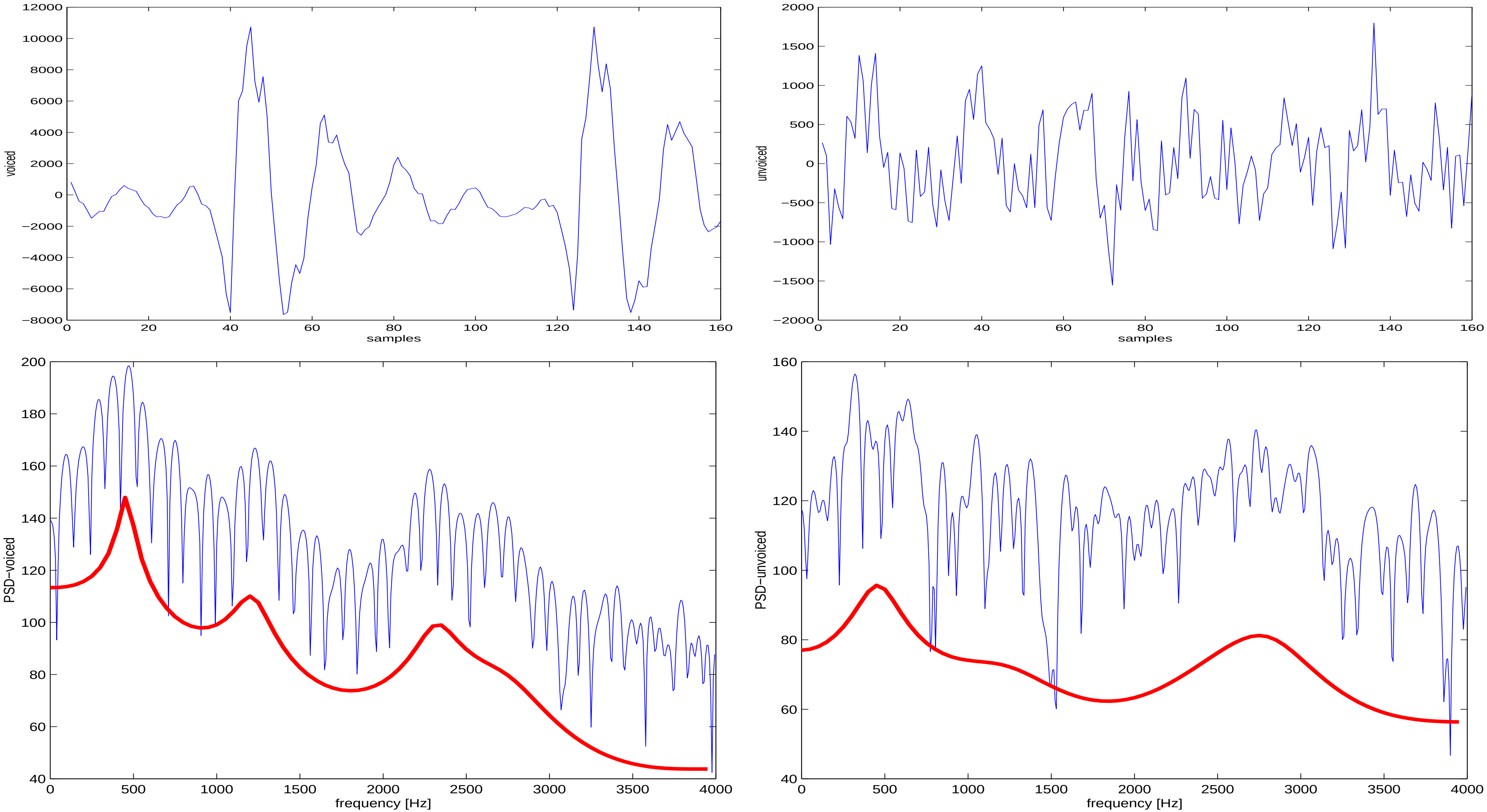
$$\hat{G}_{LPC} = \left| \frac{G}{A(z)} \right|^2_{z=e^{j2\pi f}}, \tag{28}$$

kde f je normovaná frekvence $f = \dfrac{F}{F_s}$. Po dosazení:

$$\hat{G}_{LPC} = \frac{G^2}{\left| 1 + \sum_{i=1}^{P} a_i e^{-j2\pi f i} \right|^2} \tag{29}$$

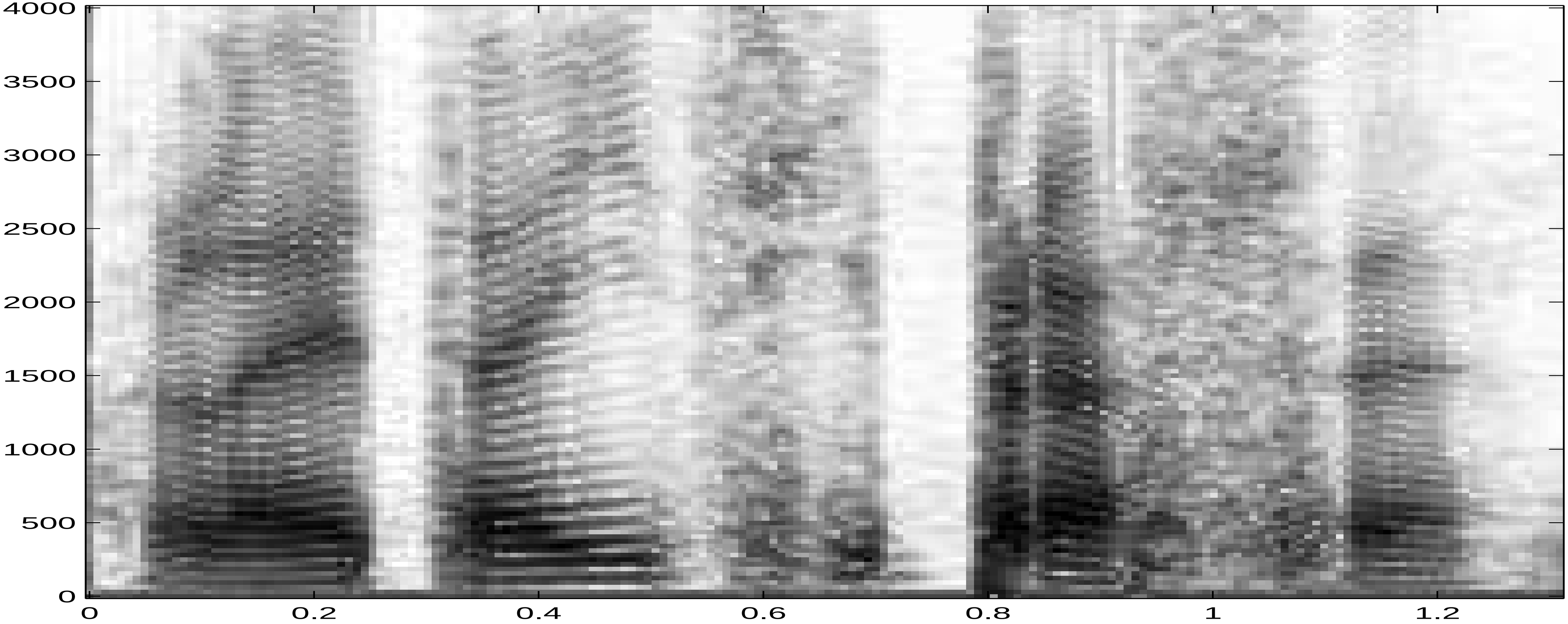
Na této spektrální hustotě výkonu se dají dobře rozeznat formanty, protože je eliminován vliv základního tónu (a tedy jemná struktura spektra, která se opakuje s F_0).

Příklad: odhad spek. hustoty získaný pomocí DFT a LPC na znělém a neznělém rámci.

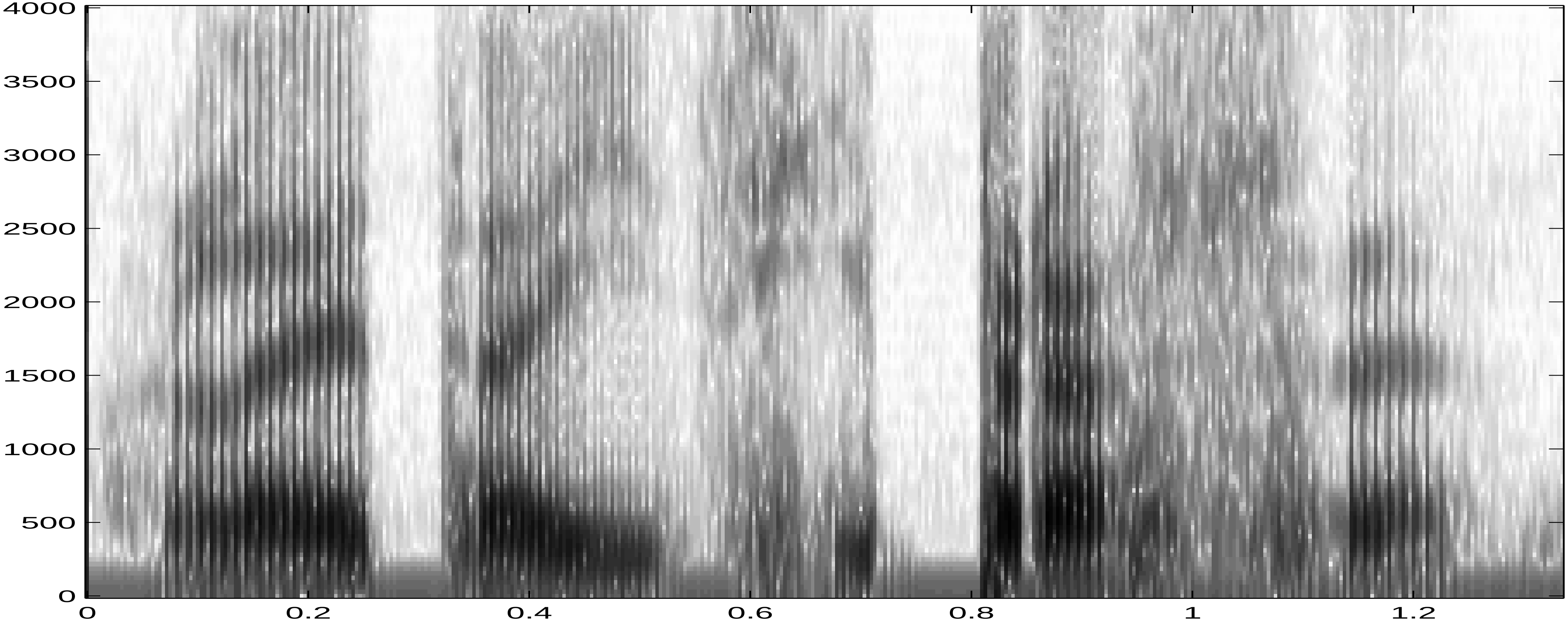


Srovnání spektrogramů

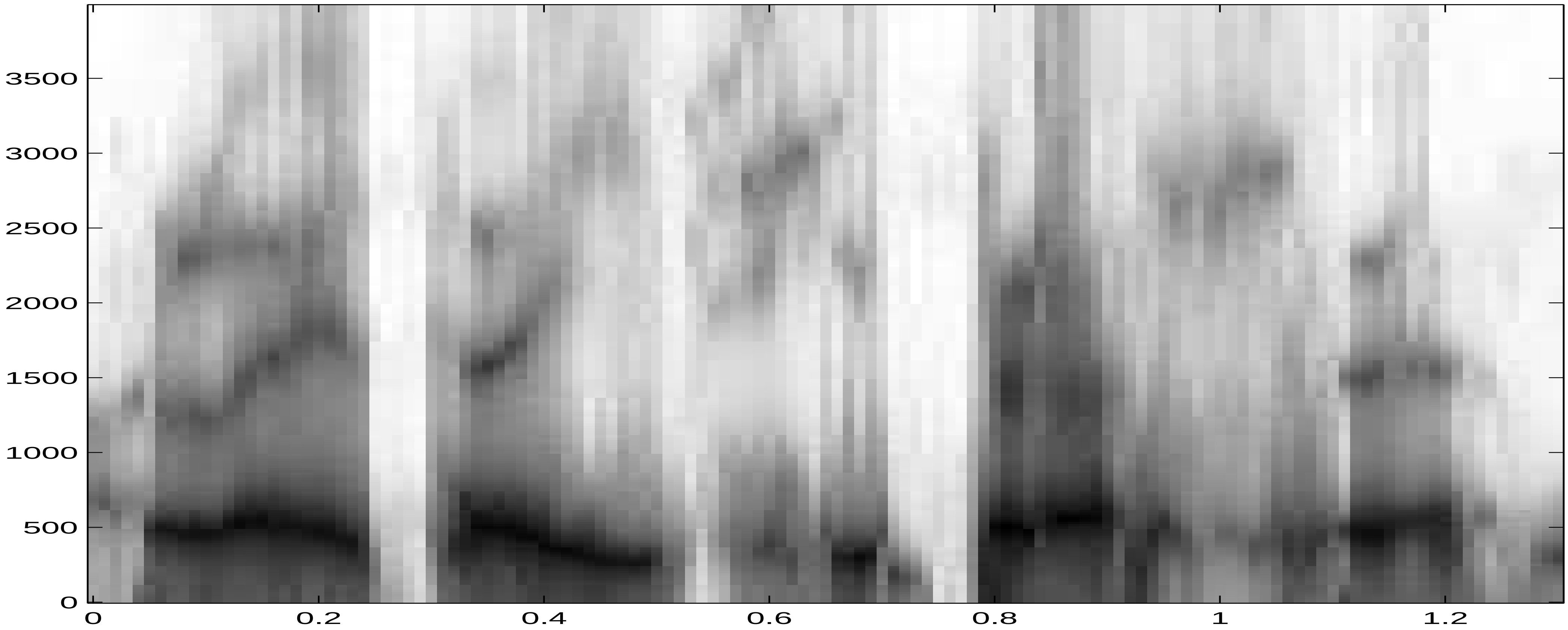
Dlouhodobý spektrogram `specgram(s,256,8000,hamming(256),200);`



Krátkodobý spektrogram `specgram(s,256,8000,hamming(50));`



LPC spektrogram



Parametry odvozené z LPC koeficientů

Proč ? Koeficienty filtru a_i jsou dobré na filtrování, ale to je tak všechno:

- Špatně se kvantují (velká citlivost stability filtru na přesnost kvantizace, nejsou omezeny: $a_i \in < -\infty, +\infty >$.)
- Jsou tvrdě korelovány – špatné pro rozpoznávání založené na HMM.
- Vzdálenost dvou vektorů koeficientů a_i nemá nic společného s podobností nebo nepodobností řečových rámců – špatné i pro rozpoznávání založené na přímém srovnávání parametrů (DTW)

$\Rightarrow$ bylo by něco lepšího ???

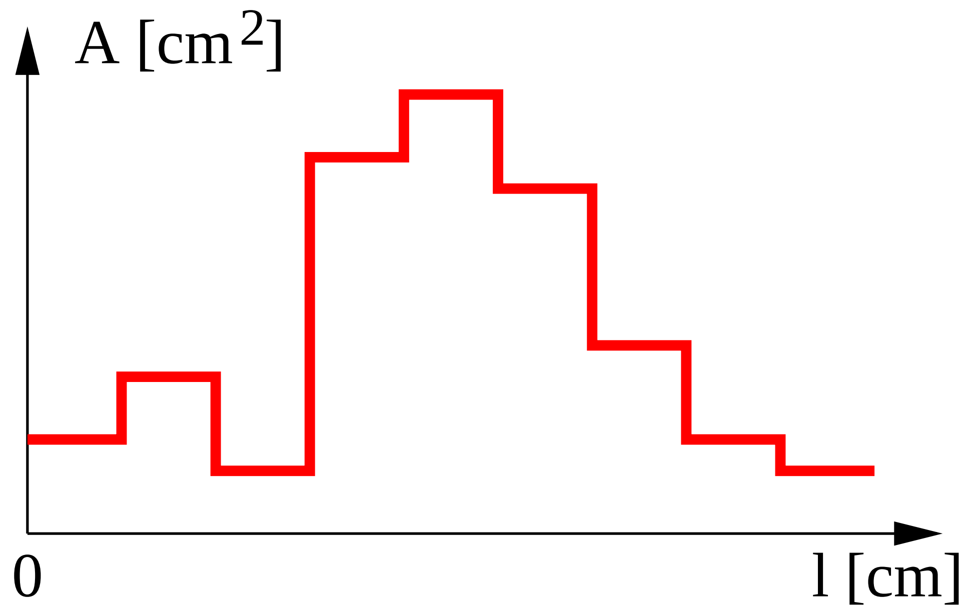
PARCOR

$$
\begin{array}{ccccc}
a_1^{(1)} & a_1^{(2)} & a_1^{(3)} & \cdots & a_1^{(P)} \\
 & a_2^{(2)} & a_2^{(3)} & \cdots & a_2^{(P)} \\
 & & a_3^{(3)} & \cdots & a_3^{(P)} \\
 & & & \ddots & \vdots \\
 & & & & a_P^{(P)}
\end{array}
$$

- Mezivýsledky Levinsona-Durbina: koeficienty $k_i = a_i^{(i)}$ se označují jako *koeficienty odrazu* nebo také *koeficienty PARCOR* (partial correlation).
- Platí pro ně: $k_i \in < -1, 1 >$. Jsou tedy vhodnější pro kódování než a_i.
- Sady koeficientů a_i a k_i se jedna na druhou dají jednoduše převést.

Válcový model hlasového traktu

hlasový trakt můžeme fyzikálně modelovat pomocí válcových sekcí o stejné délce, avšak o různých průměrech (a tím i průřezech):



Vztah válcového modelu s LPC

poměr sousedních sekcí:

$$\frac{A_{m-1}}{A_m} = \frac{1+k_m}{1-k_m} \tag{30}$$

pro $m = P, P-1, \ldots, 1$. Plocha A_P je fiktivní – skutečnou velikost neznáme, položíme tedy $A_P = 1$. Hodnoty $\frac{A_{m-1}}{A_m}$ jsou pak poměry ploch – area ratios (AR). Používanější jsou logaritmické poměry ploch – log area ratios (LAR):

$$g_m = \log \frac{A_{m-1}}{A_m} = \log \frac{1+k_m}{1-k_m} \tag{31}$$

Výhoda g_m oproti k_i je v lineární citlivosti spektra na hodnoty. Je možné použít lineární kvantifikátor hodnot g_m. U k_i je spektrum silně citlivé na hodnoty $k_i \to 0$.

LSF či LSP

Line Spectrum Frequencies (LSF) nebo také *Line Spectrum Pairs (LSP)*, jsou odvozeny z kořenů dvou polynomů:

$$
\begin{aligned}
M(z) &= A(z) - z^{-(P+1)}A(z^{-1}) \\
Q(z) &= A(z) + z^{-(P+1)}A(z^{-1}).
\end{aligned} \tag{32}
$$

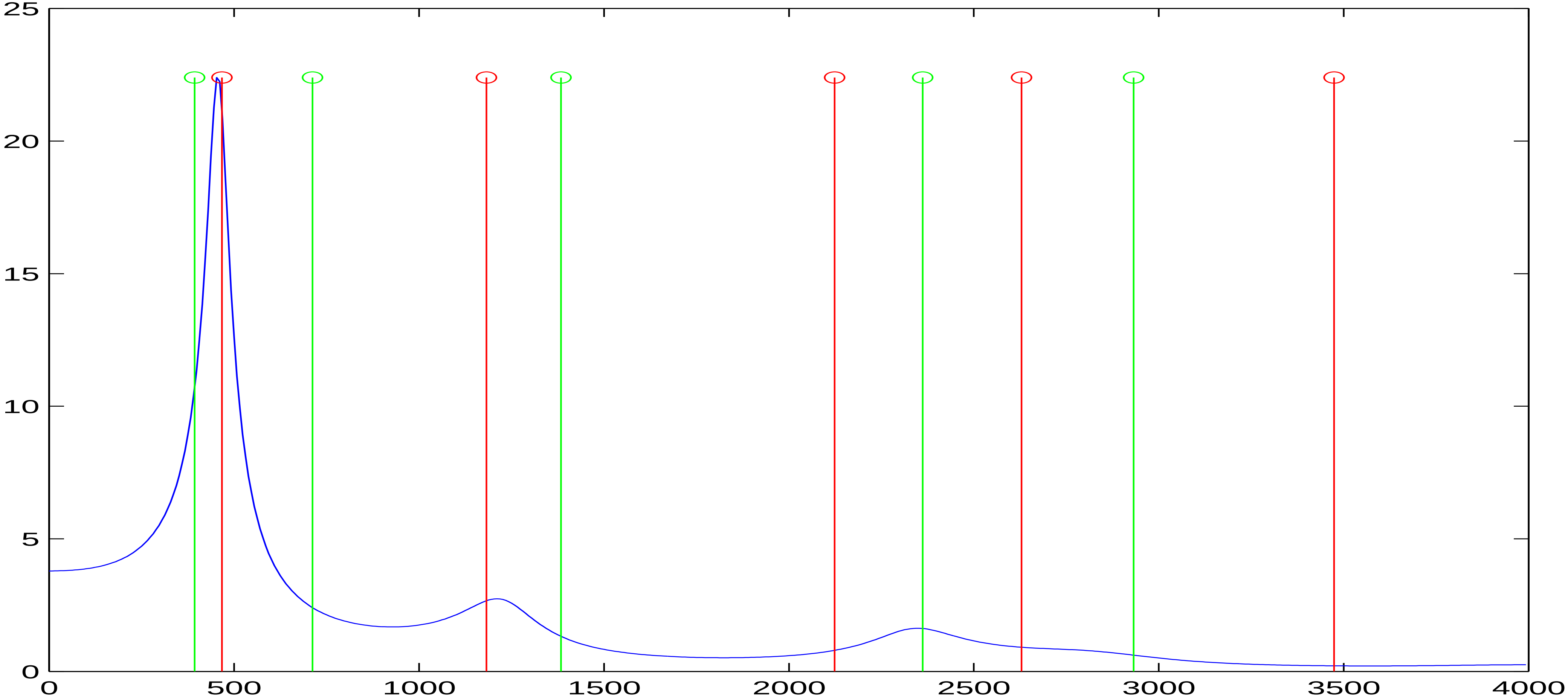
Pomocí kořenů se dají zapsat:

$$
\begin{aligned}
M(z) &= (1 - z^{-1}) \prod_{i=2,4,\ldots,P} (1 - 2z^{-1}\cos\omega_i + z^{-2}) \\
Q(z) &= (1 + z^{-1}) \prod_{i=1,3,\ldots,P-1} (1 - 2z^{-1}\cos\omega_i + z^{-2}).
\end{aligned} \tag{33}
$$

kde ω je normovaná kruhová krekvence $\omega = 2\pi f$ (f je normovaná "obyčejná" frekvence). Line spectral frequencies f_i jsou v intervalu (0,0.5) a jsou seřazeny vzestupně:

$$
0 < f_1 < f_2 < \ldots < f_{P-1} < f_P < \frac{1}{2}. \tag{34}
$$

Použijeme-li LSFs pro přenos (jsou kvantovány), můžeme v dekodéru otestovat správné dekódování tak, že zkontrolujeme jejich řazení.



LPC-cepstrum

Cepstrální koeficienty jsme počítali pomocí DFT. Je také možno určit pomocí spektrální hustoty výkonu získané odhadem pomocí LPC:

$$\hat{G}_{LPC}(f) = \left| \frac{G}{A(z)} \right|^2_{z=e^{j2\pi f}}, \tag{35}$$

kde G je gain "syntezačního" filtru a $A(z)$ je polynom řádu P. V tomto případě hovoříme o LPC-cepstru (LPCC):

$$c(n) = \mathcal{F}^{-1}[\ln \hat{G}_{LPC}(f)] \tag{36}$$

Dají se odvodit následující vlastnosti LPC-cepstrálních koeficientů:

$$c(0) = \ln G^2. \tag{37}$$

Nultý LPC-cepstrální koeficient tedy nese informaci o *energii* daného řečového rámce.

Další koeficienty lze vypočítat z LPC koeficientů pomocí rekurentních vztahů:

$$
\begin{aligned}
c(n) &= -a_n - \frac{1}{n}\sum_{k=1}^{n-1} k c_k a_{n-k} \quad \text{pro} \quad 1 \leq n \leq P \\
c(n) &= -\frac{1}{n}\sum_{k=1}^{n-1} k c_k a_{n-k} \quad \text{pro} \quad n > P
\end{aligned}
\tag{38}
$$

$\Rightarrow$ velmi jednoduchý převod.

Použití LPCC koeficientů

- LPCC koeficienty jsou jednou z parametrisací používaných v *rozpoznávačích řeči*. Jejich výhodou je, že jednotlivé koeficienty jsou méně korelovány než například LPC koeficienty a_i, v rozpoznávačích postavených na skrytých Markovových modelech (hidden Markov models – HMM) se obejdeme bez plných kovariančních matic $\mathbf{\Sigma}$ a vystačíme s vektory rozptylů. Více v kapitole o rozpozávávání pomocí HMM.
- pomocí dvou sad LPCC koeficientů můžeme jednoduše spočítat logaritmickou spektrální vzdálenost (logarithmic spectral distance) mezi dvěma řečovými rámci (nepříjemná definice s integrálem přejde na prostou Euklidovu vzdálenost dvou vektorů LPCC).